Modelling of wind power forecasting errors based on kernel recursive least-squares method

Man XU¹, Zongxiang LU¹, Ying QIAO¹, Yong MIN¹

Abstract Forecasting error amending is a universal solution to improve short-term wind power forecasting accuracy no matter what specific forecasting algorithms are applied. The error correction model should be presented considering not only the nonlinear and non-stationary characteristics of forecasting errors but also the field application adaptability problems. The kernel recursive least-squares (KRLS) model is introduced to meet the requirements of online error correction. An iterative error modification approach is designed in this paper to yield the potential benefits of statistical models, including a set of error forecasting models. The teleconnection in forecasting errors from aggregated wind farms serves as the physical background to choose the hybrid regression variables. A case study based on field data is found to validate the properties of the proposed approach. The results show that our approach could effectively extend the modifying horizon of statistical models and has a better performance

than the traditional linear method for amending short-term forecasts.

Keywords Forecasting error amending, Kernel recursive least-squares (KRLS), Spatial and temporal teleconnection, Wind power forecast

1 Introduction

After annually doubling the total installed capacity of wind turbines from 2006 to 2009, the Chinese wind power industry has entered a stable phase of development with a cumulative installed capacity of 145362 MW until the end of 2015 [1]. Wind power forecasting is a basic support technique for wind power development, which is valuable for the efficiency of energy utilization and power flow control of wind farms [2, 3]. Relevant standards and codes have stressed the application of wind power forecasting systems to mitigate the adverse effects of wind fluctuations on the safe and stable operation of power system [4, 5]. Many forecasting systems have been developed with different operation schemes, which are adaptive to field conditions [6, 7]. For the past couple of years, wind power forecasting in China has grown from zeros to a point where almost all of the large wind farms currently have at least one forecasting system operating for the daily business routine.

The typical schemes for short-term wind power forecasting (with several hours to day-ahead forecasting horizons) usually involve the physical and/or statistical approach of wind speed prediction, and the power curve model, which can convert the results from wind speed to wind power. Currently, the field forecasting accuracy in China can barely meet the expectations of users, which may be caused by numerical whether prediction (NWP)

CrossCheck date: 26 October 2016

Received: 30 December 2015 / Accepted: 27 October 2016 / Published online: 10 January 2017

© The Author(s) 2017. This article is published with open access at Springerlink.com

✉ Man XU
carriethu@foxmail.com

Zongxiang LU
luzongxiang98@tsinghua.edu.cn

Ying QIAO
qiaoying@tsinghua.edu.cn

Yong MIN
minyong@tsinghua.edu.cn

¹ State Key Lab of Power Systems, Department of Electrical Engineering, Tsinghua University, Haidian District, Beijing 100084, China

STATE GRID

STATE GRID ELECTRIC POWER RESEARCH INSTITUTE

precision, data quality, etc. Modeling forecasting errors to amend the results is a universal way to remedy the defects of prediction procedures, independent of the specific forecasting method used. Such an approach can be easily applied to improve forecasting accuracy effectively and automatically for commercial operating forecasting systems, which may cover dozens of wind farms and involve composite forecasting models.

Some studies have learned from the model output statistic (MOS) procedure of NWP to amend wind power forecasting errors, training the regression model based on the historical records of forecasting results and actual power output [8, 9]. Most of these studies applied linear models derived from the MOS procedure and originally designed for linear variables. Reference [10] indicates that the nonlinear features of forecasting errors are mainly derived from the typical nonlinear procedure of forecasting systems, i.e., the power curve model, which ideally meets Betz' law, including the cube of the wind speed. Therefore, the wind power and its forecasting error sequence reflect nonlinear characteristics after the energy conversion procedure, though the model input variables, such as wind speed, may be linear. Over all, linear models contradict the nonlinear forecasting error sequence to some extent and have some limitations to characterize the relationship between the lagged variables.

References [11, 12] applied back propagation (BP) neural networks and support vector machine (SVM) to learn the relationship between historical forecasting errors and other variables. Such intelligent learning methods use kernel function as the basis, which could extend the linear algorithms to the nonlinear relationships of variables. In this way, the fast and simple calculation steps of linear methods are retained, and the nonlinear features can be captured by the kernel-based algorithms. The basic idea is to substitute the inner product in the high-dimensional space by the kernel function value of the input vectors to avoid complex calculation procedures of the high-dimensional vectors, i.e., the kernel trick.

Owing to the stochastic nature of the process, the forecasting accuracy of the target wind farm changes with the varying situations. This results in the fluctuant and non-stationary features of forecasting errors. Several publications detect the fluctuation characteristics of historical data to capture future trends of the errors [13, 14]. Reference [13] analyzes the statistics of the fluctuation and amplitude of wind power to establish the error estimation model. Reference [14] studies the numerical characteristics of the "recent forecasting errors" to predict the error of the next moment.

The non-stationary features denote that the mean and variance of the series is time-varying. For online operation, the model should be adjusted with the updated samples at each time stamp, so that it can reflect the latest changes of

wind resource, turbines and other elements. Due to the practical constraints, the computational time of updating models should not increase significantly with the growing sample size. Thus, the challenge for the kernel-based intelligent methods is how to maintain a reasonable calculation speed with the growing sample size, in order to meet the efficiency requirement of online rolling modelling. Recursive least squares algorithms could ensure that the computational complexity introduced by a new sample is irrelevant to the sample updated time, which meets the demand of online applications [15].

The kernel recursive least-squares (KRLS) algorithm is a combination of the kernel method and recursive least squares, and has a remarkable computational time-saving effect based on its sparsity solution [16]. KRLS has been widely applied to nonlinear signal processing [17–19], and a few explorations have also been performed for wind power forecasting. Some work discussed how to improve the accuracy of very short-term wind speed forecasting using KRLS [20, 21].

In this paper, a new error-amending approach is proposed to yield the potential benefit of statistical methods, consisting of several error forecasting models, which are set in an iterative way. We introduce the KRLS method to model the forecasting errors, which is adaptive to the non-linear and non-stationary characteristics of the errors. The input variables are selected by the teleconnection analysis of forecasting errors from neighboring wind farms. This can help capture the fluctuation characteristics in view of the spatial and temporal correlations in forecasting errors, contributing to the error prediction. The presented approach can be part of the aggregating wind power forecasting platform, which breaks down the information barriers among single wind farm systems and enables our approach to have a better performance in improving forecasting accuracy.

The rest of the paper is structured as follows. Section 2 gives a brief introduction on how to extend the general regression model to the KRLS one. The design of the error amending approach and the illustration of the teleconnection of forecasting errors among neighboring wind farms are presented in Section 3. The performance of the method for different horizons of wind power forecasts is demonstrated in Section 4, based on field data from an aggregated forecasting platform in China. Section 5 gives the conclusion.

2 KRLS method

2.1 General regression form

Generally, for a set of multi-input-single-output (MISO) samples, $[x_i, y_i]$, $i = 1, 2, \dots, N$, the regression model $f(\cdot)$ can be written as:

$$y_i = f(\mathbf{x}_i) + \varepsilon_i \quad (1)$$

where y_i is the response variable at time i ; $\mathbf{x}_i = [x_i^1, x_i^2, \dots, x_i^l]^T$ is the vector of explanatory variables, and ε_i is the white noise. For wind power forecasting error modeling, y_i is the forecasting error at time i , and $\mathbf{x}_i$ is the l -dimensional vector of relevant regression variables, e.g., auto-regression variables $y_{i-\tau}, y_{i-\tau-1}, \dots, y_{i-\tau-d}$, with τ as the modifying horizon and d as the time lag, or hybrid regression variables composed of forecasting errors from neighboring wind farms.

The least-squares algorithm is used to find the estimation $\hat{f}(\cdot)$ of $f(\cdot)$ by minimizing the sum of squared residuals:

$$J = \sum_{i=1}^t (y_i - \hat{f}(\mathbf{x}_i))^2 \quad (2)$$

By multivariate linear model (MLR), $f(\cdot)$ can be expressed as:

$$y_i = \mathbf{b}^T \mathbf{x}_i + \varepsilon_i \quad (3)$$

where $\mathbf{b}$ is the weight vector, and $\mathbf{b} \in \mathbf{R}^{l \times 1}$. At time t , the cost function is:

$$J(\mathbf{b}) = \sum_{i=1}^t (y_i - \mathbf{b}^T \mathbf{x}_i)^2 = |\mathbf{Y}_t - \mathbf{X}_t \mathbf{b}|^2 \quad (4)$$

where $\mathbf{Y}_t = [y_1, y_2, \dots, y_t]^T$ and $\mathbf{X}_t = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t]^T$.

2.2 KRLS model and its recursive formulation

The response variable x is mapped into a high-dimensional space by a fixed finite map ϕ . Thus, (3) and (4) are reformulated as:

$$\hat{f}(\mathbf{x}) = \langle \mathbf{w}, \phi(\mathbf{x}) \rangle \quad (5)$$

$$J(\mathbf{w}) = \sum_{i=1}^t (y_i - \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle)^2 = |\mathbf{Y}_t - \Phi_t \mathbf{w}|^2 \quad (6)$$

where $\langle \cdot, \cdot \rangle$ means the inner product; $\mathbf{w}$ is the weight vector with the same dimension of ϕ and the map matrix $\Phi_t = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_t)]^T$. Our aim is to have:

$$\mathbf{w}_t = \arg \min_{\mathbf{w}} |\mathbf{Y}_t - \Phi_t \mathbf{w}|^2 \quad (7)$$

which is difficult to calculate in the high-dimensional space and can be expressed as:

$$\mathbf{w}_t = \sum_{i=1}^t q_i \phi(\mathbf{x}_i) = \Phi_t^T \mathbf{q} \quad (8)$$

where $\mathbf{q} = [q_1, q_2, \dots, q_t]^T$ is the coefficient vector of $\phi(\mathbf{x})$ for $\mathbf{w}_t$. Plugging this into (6) and substituting the inner product by kernel function values, we obtain:

$$J(\mathbf{q}) = |\mathbf{Y}_t - \mathbf{K}_t \mathbf{q}|^2 \quad (9)$$

where $\mathbf{K}_t$ is the kernel matrix with dimension $t \times t$, $\mathbf{K}_t(i, j) = k(\mathbf{x}_i, \mathbf{x}_j)$, $i, j = 1, 2, \dots, t$ and $k(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function, which is chosen as the Gaussian kernel in this work:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\eta^2\right) \quad (10)$$

where η is the bandwidth. The Gaussian kernel is a measure of the similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$, which is in accordance with the analysis of the correlations in forecasting errors. Studies about time series forecasting show that the Gaussian kernel could have a better performance than other functions, such as the polynomial and triangular kernels, when priori knowledge lacks [16, 22]. There are many researches on kernel methods about discussing the selection and construction of kernel functions, which continues putting forward more complicated kernels [23, 24]. Although this is not the focus of this work, we may use the relative achievements to enhance the performance of the error-amending model in future.

The classic least-squares algorithms have the dimension of $\mathbf{K}_t$ be equal to the number of samples. As the sample size increases, the calculation time increases dramatically and over fitting may occur. We employ the sparse method in [16] to solve the problems, the basic idea of which is to make use of the selected sample dictionary $D_t = \{\tilde{\mathbf{x}}_i\}$, $i = 1, 2, \dots, m_t$, $m_t \ll t$, instead of the whole dataset $\{\mathbf{x}_i\}$, $i = 1, 2, \dots, t$. Accordingly, $\mathbf{K}_t$ is substituted by the selected kernel matrix $\tilde{\mathbf{K}}_t$ with dimension $m_t \times m_t$. Thus, the amount of computation to solve $\mathbf{q}_t$ can be reduced greatly. The recursive formulations are listed as follows.

- 1) Initialization: $k_{1,1} = k(\mathbf{x}_1, \mathbf{x}_1)$, $\tilde{\mathbf{K}}_1 = k_{1,1}$, $\tilde{\mathbf{K}}_1^{-1} = 1/k_{1,1}$, $\tilde{\mathbf{q}}_1 = y_1/k_{1,1}$, the optimal expansion coefficient matrix $\mathbf{A}_1 = 1$, $\mathbf{D}_1 = \{\mathbf{x}_1\}$, and set the sparse parameter v .
- 2) At time t , the regression coefficients is updated as follows.
 - a) Calculate the kernel vector $\tilde{\mathbf{k}}_{t-1,t}, \tilde{\mathbf{k}}_{t-1,t} \in \mathbf{R}^{m_{t-1} \times 1}$ by $\tilde{\mathbf{k}}_{t-1,t}(i) = k(\tilde{\mathbf{x}}_i, \mathbf{x}_t)$, $i = 1, 2, \dots, m_{t-1}$;
 - b) Define $\mathbf{s}_t = \tilde{\mathbf{K}}_{t-1}^{-1} \tilde{\mathbf{k}}_{t-1,t}$;
 - c) Calculate the residual $\varepsilon_t = k_{t,t} - \tilde{\mathbf{k}}_{t-1,t}^T \mathbf{s}_t$.

If $\varepsilon_t > v$, then the new sample $\mathbf{x}_t$ is considered to be approximately linearly independent with the combination of D_{t-1} , and should be added to the dictionary. The updated steps are:

$$D_t = D_{t-1} \cup \{\mathbf{x}_t\} \quad (11)$$

$$\tilde{\mathbf{K}}_t^{-1} = \frac{1}{\varepsilon_t} \begin{bmatrix} \varepsilon_t \tilde{\mathbf{K}}_{t-1}^{-1} + \mathbf{s}_t \mathbf{s}_t^T & -\mathbf{s}_t \\ -\mathbf{s}_t^T & 1 \end{bmatrix} \quad (12)$$

$$\mathbf{A}_t = \begin{bmatrix} \mathbf{A}_{t-1} & 0 \\ \mathbf{0}^T & 1 \end{bmatrix} \quad (13)$$

$$\tilde{\mathbf{q}}_t = \begin{bmatrix} \tilde{\mathbf{q}}_{t-1} - \frac{\mathbf{s}_t}{\varepsilon_t} (y_t - \tilde{\mathbf{k}}_{t-1,t}^T \tilde{\mathbf{q}}_{t-1}) \\ \frac{1}{\varepsilon_t} (y_t - \tilde{\mathbf{k}}_{t-1,t}^T \tilde{\mathbf{q}}_{t-1}) \end{bmatrix} \quad (14)$$

Else, if $\varepsilon_t \leq v$, then $\mathbf{x}_t$ could be approximately expressed by D_{t-1} . Therefore, $D_t = D_{t-1}$, the updated steps are:

$$\mathbf{A}_t = \mathbf{A}_{t-1} - \frac{\mathbf{A}_{t-1} \mathbf{s}_t \mathbf{s}_t^T \mathbf{A}_{t-1}}{1 + \mathbf{s}_t^T \mathbf{A}_{t-1} \mathbf{s}_t} \quad (15)$$

$$\tilde{\mathbf{q}}_t = \tilde{\mathbf{q}}_{t-1} + \tilde{\mathbf{K}}_t^{-1} \mathbf{r}_t (y_t - \tilde{\mathbf{k}}_{t-1,t}^T \tilde{\mathbf{q}}_{t-1}) \quad (16)$$

where $\mathbf{r}_t$ is defined by $\mathbf{r}_t = \mathbf{A}_{t-1} \mathbf{s}_t / (1 + \mathbf{s}_t^T \mathbf{A}_{t-1} \mathbf{s}_t)$.

3) Given x_i , the corresponding estimation is:

$$\hat{y}_i = \tilde{\mathbf{k}}_{t,i}^T \tilde{\mathbf{q}}_t \quad (17)$$

where $\tilde{\mathbf{k}}_{t,i}(j) = k(\tilde{\mathbf{x}}_j, \mathbf{x}_i)$, $j = 1, 2, \dots, m_t$.

3 Error amending of wind power forecasts

3.1 Amending approach

The historical samples are supposed to imply the properties of wind power forecasting errors. The designed approach is to modify the original forecasts by error forecasting model, which is trained and adjusted in the historical dataset. The process can be explained as follows.

Step 1: Partition the historical dataset into two time-lapse stages: DB^- and DB^+ , and the forecasting error of DB^- is:

$$x_t = p_t^F - p_t \quad t \in DB^- \quad (18)$$

where p_t^F is the forecasting power at time t and p_t is the actual power.

Step 2: Train the first order error forecasting model f_1 for $\{x_t, t \in DB^-\}$ and apply it to modify the forecasts of DB^+ :

$$p_t^{F,1} = p_t^F - x_t^{f_1} \quad t \in DB^+ \quad (19)$$

where $x_t^{f_1}$ is the forecast of x_t .

Step 3: Test if either of the terminal conditions is valid:

1) the modified forecasting error $x_{t,1} = p_t^{F,1} - p_t$, $t \in DB^+$ meets the required precision; 2) the modification order has reached the scheduled number. If so, go to Step 6; otherwise, go to Step 4.

Step 4: Train the second order error forecasting model f_2 for $x_{t,1} = p_t^{F,1} - p_t$, $t \in DB^-$ and apply it to modify the forecasts of DB^+ :

$$p_t^{F,2} = p_t^F - x_t^{f_1} - x_{t,1}^{f_2} \quad t \in DB^+ \quad (20)$$

where $x_{t,1}^{f_2}$ is the forecast of $x_{t,1}$.

Step 5: Similar to Step 3, check to decide whether the model order should be increased again as in Step 4 or if the target is attained and we can proceed to Step 6.

Step 6: A set of modification models $\{f_i, i = 1, 2, \dots, n\}$ with the highest order n have been trained for the historical dataset $DB = DB^- + DB^+$.

Hybrid variables composed by forecasting error from the target wind farm and its neighbors are recommended by the authors for model f_1 . For model f_i , $i > 1$, auto-regression is preferred considering practical problems such as computational efficiency and storage space for online systems.

The above process is mainly used to determine the modeling order, the type of input variables and the value of key parameters for recursive methods. For non-recursive methods, the model trained in DB can be directly applied to the next time domain.

The flowchart of the iterative forecasting error modification approach is summarized in Fig. 1. The final modification for the original forecast at time t is:

$$p_t^{F,n} = p_t^F - \sum_{i=1}^n x_t^{f_i} \quad (21)$$

where n is the highest modeling order of the amending approach.

3.2 Teleconnection in forecasting errors

The spatial and temporal teleconnection in forecasting errors is the physical basis of whether the error amending approach could have a significant effect on forecasting accuracy. Due to the inertia of meteorological phenomena, the current forecasting error of an upwind wind farm may have a strong correlation with that of a downwind wind farm [25]. For meteorological studies, teleconnection means the stable correlation between synoptic processes that are away from each other spatially and temporally [26, 27]. Considering the similar physical background, we extend this concept to the stable correlation between forecasting errors of different wind farms

Let $\{x_{j,t}\}$ be the forecasting error sequence and j be the wind farm number. The teleconnection between forecasting errors of different wind farms is expressed by the cross-correlation function (CCF) as below:

Fig. 1 Flowchart of iterative forecasting error modification approach

$$r_{ij}(d) = r(x_{i,t}, x_{j,t-d}) = \frac{E[(x_{i,t} - \mu_i)(x_{j,t-d} - \mu_j)]}{\sigma_i \sigma_j} \quad (22)$$

where $r_{ij}(d)$ is the CCF value between wind farm i and j with the time lag d ; μ_i and μ_j are the means of $\{x_{i,t}\}$ and $\{x_{j,t-d}\}$ respectively; σ_i and σ_j are the corresponding standard deviation. It is of great importance for error-amending models with hybrid input variables to understand the teleconnection characteristics. After that, we can choose the proper parameters and regression variables based on the temporal and spatial correlation regularities of error fluctuation.

4 Case study

4.1 Data description

The data for the case study have been collected from an aggregated wind power forecasting platform in Jilin, China, which consists of seven wind farms, as shown in

Fig. 2. Please note that $Cap_a = 99$ MW, $Cap_b = 49.5$ MW, $Cap_c = 99$ MW, $Cap_d = 57.35$ MW, $Cap_e = 249.9$ MW, $Cap_f = 400.5$ MW, $Cap_g = 49.5$ MW, where Cap_a , Cap_b , Cap_c , Cap_d , Cap_e , Cap_f , Cap_g are the capacity of WFa to Wfg respectively. Both short-term forecasts, with forecasting horizons ranging from 1 to 24 hours ahead, and actual power data are collected and the time window covers from May, 2014 to October, 2015. The sampling interval of the forecasting and actual power data is 1 hour. The preprocessing work has removed the data that are apparently erroneous or influenced by wind power curtailment. The wind farm marked as WFa in Fig. 2 is chosen as the target wind farm, whose forecasting errors are to be amended. Figure 3 gives the auto-correlation function (ACF) of forecasting errors of WFa at different time lags. One can see that remarkable auto-correlation is seen at time lags that are less than 5 to 6 hours, which means that the forecasting error is very different from a white noise.

Figure 4 shows the CCF results of WFa and other wind farms. One can see that there are significant lag correlations between WFa and its neighbors, WFe and Wff, at time lags that are less than 10 hours. Therefore, according to the teleconnection analysis, forecasting errors from WFe and Wff are selected as the potential input variables to improve the forecasts of WFa. Note that for this case the teleconnection is easily captured because West (W) to Southwest (SW) is the prevailing wind direction of this region through 12 months, as shown in Fig. 5. Note that NNE stands for northeast by north; ENE stands for northeast by east; ESE stands for southeast by east; SSE stands for southeast by south; SSW stands for southwest by south; WSW stands for southwest by west; WNW

Fig. 2 Positions of sampling wind farms used in case study

Fig. 3 Auto-correlation of WFa (blue lines represent 95% confidence interval)

stands for northwest by west; NNW stands for northwest by north.

For cases under complex wind conditions, more attention should be paid to the influence of wind direction on the teleconnection in forecasting errors. If so, seasonal models according to different prevailing wind directions may be required.

4.2 Results

1) Modeling of wind power forecasting errors based on KRLS

Error amending for WFa is used as the example to illustrate the establishment of KRLS model, for which the modifying horizons are 1 to 24 hours, i.e., $\tau = 1, \dots, 24$.

First, to train the first order model, the response variable and explanatory variables in (1) are defined as: the

forecasting error at WFa, $y_i = x_{a,i}$, and the historical forecasting error at WFa, WFe, Wff, $\mathbf{x}_i = [x_{a,i-\tau}, \dots, x_{a,i-\tau-d_1}, x_{e,i-\tau}, \dots, x_{e,i-\tau-d_2}, x_{f,i-\tau}, \dots, x_{f,i-\tau-d_3}]^T$, where d_1 is the auto-regression step of WFa; and d_2, d_3 are the hybrid regression steps of WFe and Wff, respectively. A modification model with modeling order n ($n > 1$) is established based on the auto-regression of the forecasting error after modification with order $n - 1$.

The key parameters to be chosen include the sparse parameter v_n for each order, the bandwidth η_n , the hybrid regression step d_1, d_2, d_3 , and the auto-regression step $d_{ar,n}$. Generally, larger regression steps and smaller bandwidth parameters can help improve the modelling accuracy. When they are too large or small, however, the improvement on forecasting accuracy is very limited, resulting in a waste of computation time. Thus the optimal parameter selection principle is to check whether the mean squared error (MSE) alters significantly with the changes in parameters.

The total number of sample points after preprocessing is 4560, which are then normalized to the interval $[0, 1]$ and divided into three groups: 1) 1 to 500 points are set for modeling initialization to accumulate the sample dictionary and steady the modeling effects; 2) 501 to 2000 are for the cross validation test of parameters; 3) The remaining 2001 to 4560 points are the accuracy test samples. The final models are determined as:

- First order: $v_1 = 0.1$, $\eta_1 = 1.6$, $d_1 = 6$, $d_2 = 4$, $d_3 = 4$;
- Second order: $v_2 = 0.01$, $\eta_2 = 0.9$, $d_{ar,2} = 3$.

2) Model comparison

The proposed model is marked as $f_{K,M}^2$. To have a further look at the effect of modification, the MLR is chosen as the benchmark method, which is widely used for MOS modeling. At time t , for samples $[x_i, y_i]$, $i = 1, 2, \dots, t$, the regression model is:

$$\mathbf{Y}_t = \mathbf{X}_t' \mathbf{b}_t' + \varepsilon_t \quad (23)$$

where $\mathbf{Y}_t = [y_1, y_2, \dots, y_t]^T$; $\mathbf{X}_t' = [\mathbf{I} \quad \mathbf{X}_t]$; $\mathbf{I}$ is the unit vector; $\mathbf{X}_t = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t]^T$; $\varepsilon_t = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_t]^T$; $\mathbf{b}_t' = [b_t^0, b_t^1, \dots, b_t^l]^T$ and l is the dimension of $\mathbf{x}_i$. $\mathbf{b}_t'$ is estimated by least squares as:

$$\hat{\mathbf{b}}_t' = (\mathbf{X}_t'^T \mathbf{X}_t')^{-1} \mathbf{X}_t'^T \mathbf{Y}_t \quad (24)$$

which can then be extended to the next time domain to estimate the forecasting error.

All of the models that are compared with $f_{K,M}^2$ are summarized in Table 1, and Fig. 6 gives the forecasting

Fig. 4 CCF value of wind farm between WFa and its neighboring wind farms (blue lines represent 95% confidence interval)

Fig. 5 Rose maps of wind direction of WFa, WFe, WFF

error amending results of all of the models, which is judged by the root mean squares error (RMSE). Note that the maximum value of the horizontal coordinates is set to be 14 due to the fact that all the models are failed when the modifying horizon is longer than 14 hours. The RMSE has been calculated as a percentage of nominal capacity of the wind farm. It can be seen that the modification effect varies with changing modifying horizons and different models. With the same type of explanatory variables, all of the MLR models have a poorer performance than KRLS. When the horizon is very short, i.e., $\tau = 1$, the difference is not large. However, the performance of MLR models decreases very quickly as the horizon grows and the advantage of KRLS increases. When the modified results approach the original accuracy, the improvement will fluctuate slightly around the original one. Thus we identify the failure of the modification when the difference between the modified error and original error is less than 0.05%, the previous horizon of which is defined as the effective modifying horizon. Due to this result, the effective horizons of $f_{M,A}^1$, $f_{K,A}^1$, $f_{M,M}^1$, $f_{K,M}^1$ and $f_{K,M}^2$ are 5, 9, 7, 11, 11 hours, respectively. It can be seen that the effective horizon could be increased by applying more complicated nonlinear models and considering the teleconnection for input variables as well.

4.3 Discussion

1) Key factors of the modification effect for different horizons

To analyze the key factors of the amending effect, Fig. 7 is a bar diagram showing the improvement on RMSE by different models. When a model fails, the improvement will be set as 0. The key factors are as follows.

- Base value: $f_{M,A}^1$ is a simple auto-regression linear model, which can be regarded as the base value of the modification;
- NL + NS: The improvement by $f_{K,A}^1$ is due to its applicability to the nonlinear and non-stationary (NL + NS) characteristics of the samples;
- TC: The advantage of $f_{K,M}^1$ over $f_{K,A}^1$ is the consideration of the teleconnection (TC) in the samples;
- IT: $f_{K,M}^2$ can modify the error of forecasting error iteratively (IT) for further improvement.

It can be seen from Fig. 7 that the contribution of 'NL + NS' to forecasting accuracy grows significantly as the horizon increases from 1 to 3 hours, which gradually becomes the major influencing factor of error amending. On the contrary, the importance of base value and 'TC' that are both based on the spatial and temporal correlation of the samples gradually wear off at such horizons, which

Table 1 Models to be compared with $f_{K,M}^2$

Effective horizon	Type	Regression steps	Parameters
$f_{M,A}^1$	Auto-regression, first order MLR	6	$v_1 = 0.04, \eta_1 = 1.1$
$f_{K,A}^1$	Auto-regression, first order KRLS	6	
$f_{M,M}^1$	Hybrid regression, first order MLR	$d_1 = 6, d_2 = 3, d_3 = 2$	$v_1 = 0.1, \eta_1 = 1.6$
$f_{K,M}^1$	Hybrid regression, first order KRLS	$d_1 = 6, d_2 = 4, d_3 = 4$	

Table 2 Forecasting error statistics of $f_{K,M}^1$ and $f_{K,M}^2$

Error	Mean error (%)	Variance	RMSE (%)	Maximal absolute error (%)
e_0	-1.0567	205.8375	14.4819	71.5974
e_1^1	-0.0119	84.5192	9.2152	60.3221
e_2^1	-0.0097	75.4361	8.7524	55.7483
e_1^3	0.1470	153.4307	12.4758	63.2212
e_2^3	0.1102	147.2122	12.2067	61.8991
e_1^6	0.3246	188.1231	13.8455	66.7483
e_2^6	0.1377	185.3274	13.7045	68.6248
e_1^{10}	-0.1988	202.03	14.3374	71.8773
e_2^{10}	-0.0502	200.6527	14.2811	71.8324

Note: e_0 is the original error; e_1^1, e_2^1, e_1^3 and e_1^{10} are modified errors of $f_{K,M}^1$ at horizons 1, 3, 6, 10; e_2^1, e_2^3, e_2^6 , and e_2^{10} are modified errors of $f_{K,M}^2$ at horizons 1, 3, 6, and 10

conforms to the ACF and CCF results in Figs. 3 and 4. When the horizons continue increasing from 4 to 10 hours, the modification gradually becomes saturated with the contribution of ‘NL + NS’ decreasing and the teleconnection is the key factor to increase the effective modifying horizon of $f_{K,M}^1, f_{K,M}^2$. Note that in Fig. 7, base value, NL + NS, TC and IT are the error modifying results from, and respectively. The RMSE has been calculated as a percentage of nominal capacity of the wind farm.

Table 2 demonstrates the forecasting error statistics of $f_{K,M}^1$ and $f_{K,M}^2$ to help illustrate the effect of multi-order models, including mean error (ME), variance, RMSE and maximal absolute error. One can see that the error bias,

reflected by ME, has significantly decreased after the first order modification. Other statistical metrics can also be reduced when the horizon is short enough. The second-order model can achieve a better performance at shorter horizons, helping lower the amplitude and volatility of forecasting error. Although such a performance drops off as the horizon increases, most of its metrics are still better than those of the first order model. The principle of increasing the order of modification model is to strike a reasonable balance between improving the accuracy and model simplification. This is the reason why higher order models are not applied in this case study.

2) Modification for different error intervals

The modified RMSEs for different error intervals are summarized in Table 3, in which it can be seen that the contribution of $f_{K,M}^2$ is mainly due to its correction on the large forecasting errors, which wears off as the horizon increases. The phase error of short-term wind power forecasting is the lead or lag of the forecasts to actual power in the horizontal time axis [6]. A lot of large RMSE results are caused by such a kind of forecasting errors. At short time horizons, the phase error can be corrected to some extent by $f_{K,M}^2$ due to its consideration for the teleconnection of forecasting errors, resulting in smaller RMSE values for large error intervals.

Table 3 Modified RMSEs for different error intervals

Error	0%–5%	5%–10%	10%–20%	>20%
e_0	2.8151	7.8321	14.9698	33.8136
e_2^1	2.8145	6.5824	11.7521	16.5531
e_2^3	2.8821	7.3761	13.4592	27.2919
e_2^6	2.7861	7.6534	14.2878	31.9532
e_2^{10}	2.8567	7.8089	14.6803	33.1829

Note: Error intervals 0%–5%, 5%–10%, 10%–20%, >20% are set according to the absolute value of original error. e_2^1, e_2^3, e_2^6 , and e_2^{10} are modified errors of $f_{K,M}^2$ at horizons 1, 3, 6, and 10. The RMSE has been calculated as a percentage of nominal capacity of the wind farm

Fig. 6 Comparison of different modification models

Fig. 7 Key factors to improve forecasting accuracy at each modifying horizon

3) Computational time

A reduction in computational time for online application is ensured by the sparse feature of KRLS, which uses the selected sample dictionary D_t for training instead of the whole dataset. The best combination of forecasting accuracy and computational efficiency is achieved by adjusting the sparse parameter ν . For example, Fig. 8 shows the results of ν_1 for f_{KM}^2 . Note that the calculation involved is based on the computing environment of MATLAB R2013a and Inter Core i7-3520M CPU of 2.90 GHz with an 8.00

Fig. 8 Computational efficiency with different values of sparse parameter

GB RAM. With decreasing values of ν_1 , the elapsed time keeps growing, while the changes in RMSE can be divided to three stages.

Stage 1: When $\nu_1 > 0.1$, RMSE decreases with the reducing ν_1 , because the size of D_t is not large enough to ensure better forecasting accuracy;

Stage 2: When $0.05 < \nu_1 < 0.1$, RMSE remains approximately the same level;

Stage 3: When $\nu_1 < 0.05$, RMSE increases dramatically with the reducing ν_1 , due to the over-fitting effect with a redundant sample dictionary.

Thus, we choose 0.1 as the best value for ν_1 . The principle to achieve the best combination is to have the best level of RMSE (e.g., no more than the minimum value of $\text{RMSE} + 0.02\%$) with the elapsed time as less as possible.

5 Conclusion

This paper presents a novel error-amending approach for short-term wind power forecasting systems to improve accuracy, which is independent of specific forecasting algorithms. The approach consists of several error-amending models based on the KRLS method to reduce the error iteratively. The teleconnection in forecasting errors is probed to help establish the hybrid regression models for aggregated wind power forecasting systems. Compared with the linear methods derived from the MOS procedure of NWP, the proposed approach is more adaptive to the nonlinear and non-stationary characteristics of the forecasting errors, and the improvement effect is verified by the results of the case study.

In addition to reducing forecasting error, it is of importance to learn the mechanism of the accuracy improvement for different modifying horizons. The simulation results show that the nonlinear and non-stationary characteristics of the samples gradually become the primary factor for short modifying horizons that are less than 4 hours, and the teleconnection is the key factor for long-horizon modification. The multi-order models contribute to maximizing the effect of statistical methods at different horizons. The simulation also confirms the especial modification effect of our approach for large forecasting errors, which may help reduce the phase error for forecasting systems.

The proposed approach has merits for online application as well. The deficiency of weak generalization ability for many kernel-based intelligent algorithms such as support vector machines requires large datasets and long training times to ensure the modeling performance, which is computationally expensive and ineffective for online rolling situations. Such a problem can be significantly mitigated by the recursive and sparse features of KRLS. Simulation

results show that the elapsed time is distinctly saved on the premise of modification effect by choosing a proper sparse parameter.

Another concern for online application of our approach is an open forecasting platform to support the data transfer among different wind farms, which can provide a solid foundation for improving the forecasting accuracy based on errors from neighboring wind farms. The hardware and software requirements for such a platform has been discussed in our recent work [28], which is an internet based platform with the forecasting server implemented at the service provider side. The target of the platform is to intelligently and flexibly deal with the teleconnection between wind farms, which may learn from the structure of multi-agent systems in future [29, 30].

Acknowledgements This work was partly supported by National Natural Science Foundation of China (No. 51190101) and science and technology project of State Grid, Research on the combined planning method for renewable power base based on multi-dimensional characteristics of wind and solar energy. The authors would like to thank Tsingsoft Innovation Technology Co. Ltd for providing the wind power data. We also thank Professor Pierre Pinson for some innovative ideas.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- [1] Fried L, Qiao L, Sawyer S et al (2015) Global Wind Energy Council (GWEC). Global wind report: annual market update. <http://www.gwec.net/publications/global-wind-report-2/global-wind-report-2015-annual-market-update/#>
- [2] Sun QY, Zhou JG, Guerrero JM et al (2015) Hybrid three-phase/single-phase microgrid architecture with power management capabilities. *IEEE Trans Power Electron* 30(10):5964–5977
- [3] Wang CS, Li XL, Guo L et al (2014) A nonlinear-disturbance-observer-based DC-bus voltage control for a hybrid AC/DC microgrid. *IEEE Trans Power Electron* 29(11):6162–6177
- [4] GB/T 19963—2011 (2011) Technical rule for connecting wind farm to power system
- [5] NB/T 31046-2013 (2013) Function specification of wind power forecasting system. People's Republic of China energy industry standards
- [6] Giebel G, Brownsword R, Kariniotakis G et al (2011) The state of the art in short-term prediction of wind power: a literature overview, 2nd edn. ANEMOS.plus, Paris
- [7] Yang M, Lin Y, Zhu SM et al (2015) Multi-dimensional scenario forecast for generation of multiple wind farms. *J Mod Power Syst Clean Energy* 3(3):361–370. doi:10.1007/s40565-015-0110-6
- [8] Lazić L, Pejanović G, Živković M et al (2014) Improved wind forecasts for wind power generation using the Eta model and mos (model output statistics) method. *Energy* 73:567–574
- [9] Yang HY, Feng SL, Wang B et al (2013) Study of the MOS method based on linear regression for wind power prediction. *Proc CSU-EPSCA* 25(4):14–17
- [10] Pinson P (2006) Estimation of the uncertainty in wind power forecasting. Ph D Thesis. Ecole des Mines de Paris, Paris, France
- [11] Mao MQ, Cao Y, Chang LC (2013) Improved fast short-term wind power prediction model based on superposition of predicted error. In: Proceedings of the 4th IEEE international symposium on power electronics for distributed generation systems (PEDG'13), Rogers, AR, USA, 8–11 Jul 2013, 6 pp
- [12] Li X, Wang X, Zheng YH et al (2015) Short-term wind load forecasting based on improved LSSVM and error forecasting correction. *Power Syst Prot Contr* 43(11):63–69
- [13] Zhang KF, Yang GQ, Chen HY et al (2014) An estimation method for wind power forecast errors based on numerical feature extraction. *Autom Electr Power Syst* 38(16):22–27. doi:10.7500/AEPS20140113008
- [14] Ye Y, Ren C, Zhao YN et al (2016) Stratification analysis approach of numerical characteristics for ultra-short-term wind power forecasting error. *Proc CSEE* 36(3):692–700
- [15] Madsen H, Holst J (2000) Modelling non-linear and non-stationary time series. Technical University of Denmark, Copenhagen
- [16] Engel Y, Mannor S, Meir R (2004) The kernel recursive least-squares algorithm. *IEEE Trans Signal Process* 52(8):2275–2285
- [17] Tan FX, Liu ZR, Yang TT (2015) Optimal control of nonaffine nonlinear discrete-time systems using kernel-based adaptive dynamic programming. In: Proceedings of the 5th international conference on information science and technology (ICIST'15), Changsha, China, 24–26 Apr 2015, pp 588–593
- [18] Vaerenbergh SV, Via J, Santamaria I (2007) Nonlinear system identification using a new sliding-window kernel RLS algorithm. *J Commun* 2(3):1–8
- [19] Zheng YF, Wang SY, Feng JC et al (2016) A modified quantized kernel least mean square algorithm for prediction of chaotic time series. *Digit Signal Process* 48:130–136
- [20] Dowell J, Weiss S, Infield D (2015) Kernel methods for short-term spatio-temporal wind prediction. In: Proceedings of the 2015 IEEE power and energy society general meeting, Denver, CO, USA, 26–30 Jul 2015, 5 pp
- [21] Kuh A, Mandic D (2009) Applications of complex augmented kernels to wind profile prediction. In: Proceedings of the 2009 IEEE international conference on acoustics, speech and signal processing (ICASSP'09), Taipei, China, 19–24 Apr 2009, pp 3481–3584
- [22] Smola AJ (1998) Learning with kernels. Technical University of Berlin, Berlin
- [23] Wang HZ, Yu JS (2006) Study on the kernel-based methods and its model selection. *J South Yangtze Univ (Nat Sci Ed)* 5(4):500–504
- [24] Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge
- [25] Tastu J, Pinson P, Kotwa E et al (2011) Spatio-temporal analysis and modeling of short-term wind power forecast errors. *Wind Energy* 14(1):43–60
- [26] Zhao QL, Li CY (2012) Dipole mode of the sea level pressure anomalies in Asia-Pacific region and the relation to winter climate anomaly in China. *Clim Environ Res* 17(1):1–12
- [27] Wallace JM, Gutzler DS (1981) Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon Weather Rev* 109(4):784–812
- [28] Lu ZX, Xu M, Qiao Y et al (2016) New Internet based operation pattern design of wind power forecasting system. *Power Syst Technol* 40(1):125–131

- [29] Sun QY, Han RK, Zhang HG et al (2015) A multiagent-based consensus algorithm for distributed coordinated control of distributed generators in the energy internet. *IEEE Trans Smart Grid* 6(6):3006–3019
- [30] Ren FH, Zhang MJ, Soetanto D et al (2012) Conceptual design of a multi-agent system for interconnected power systems restoration. *IEEE Trans Power Syst* 27(2):732–740

Man XU received the B.S. degree in electrical engineering from Tsinghua University. She is now a PHD candidate in electrical engineering at State Key Laboratory of Power System, Tsinghua University, Beijing, China. Her research interests include wind power forecast and data processing.

Zongxiang LU received the B.S. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1998 and 2002, respectively. He is now an associate Professor of Electrical Engineering at Tsinghua University, Beijing, China. His research interests include power system reliability, renewable energy and

microgrid, and large scale wind power integration. He has been the PI of more than 30 academic and industrial projects. He is the author of 2 books on power system reliability and wind power integration. He was awarded F5000 influential papers and top articles in China in 2007.

Ying QIAO received the B.S. and Ph.D. degrees in electrical engineering from Shanghai Jiaotong University, Shanghai, and Tsinghua University, Beijing, China, in 2002 and 2008, respectively. She is now an associate Professor of Electrical Engineering at Tsinghua University, Beijing, China, where she has been employed since 2010. Her research interests include renewable energy and power system security and control.

Yong MIN received the B.Sc. degree and the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 1984 and 1990, respectively. He is currently a Professor with the Department of Electrical Engineering, Tsinghua University. His research interests are in power system stability and control. Prof. Min is a Fellow of the IET.